DOCUMENT RESUME

ED 343 691 PS 020 407

AUTHOR Halford, Graeme S.
TITLE Working Memory and Reasoning: The Processing Loads Imposed by Analogies.
SPONS AGENCY Australian Research Grants Scheme.
PUB DATE Jul 91
NOTE 31p.; Paper presented at the Biennial Meeting of the International Society for the Study of Behavioral Development (11th, Minneapolis, MN, July 3-7, 1991).
PUB TYPE Speeches/Conference Papers (150) -- Information Analyses (070)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Adults; Age Differences; Children; *Cognitive Development; *Cognitive Mapping; Cognitive Structures; Developmental Stages; *Encoding (Psychology); *Learning Modalities; *Memory; Metacognition; *Thinking Skills

ABSTRACT

The proposals concerning working memory outlined in this paper involve the architecture of working memory, the reasoning mechanisms that draw on it, and the ways in which working memory may develop with age. Ways of assessing task demands and children's working memory capacities are also considered. It is noted that there is long-standing evidence that much human reasoning is analogical. Recently developed models have shown that analogies can be conceptualized as a mapping from a base structure to a target structure. This paper begins with a review of the concepts of short-term and primary memory, followed by a literature review. Discussion then turns to the question of whether working memory increases or remains constant with age. It is maintained that the lack of adequate measures of working memory capacity makes it impossible to resolve the question. However, when the question is analyzed in terms of neural net architecture, the hypothesis that overall capacity remains constant but representations become more differentiated with age appears theoretically and empirically reasonable. If the hypothesis were valid, it would mean that the ability to map structures increases with age, and the development of reasoning capacity has a maturational component. Contains 35 references. (LB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED343691

Working Memory and Reasoning: The Processing Loads Imposed by Analogies

Graeme S. Halford

University of Queensland

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
G.S. Halford
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Running head: PDP and Working Memory

Paper presented to Eleventh Biennial Meeting of the International Society for the Study of Behavioral Development, Minneapolis, Minnesota, USA, July 3-7, 1991.

This work was supported by grants from the Australian Research Grants Scheme. Author's address: Graeme Halford, Psychology Department, University of Queensland, 4072, AUSTRALIA, AIR MAIL. Fax: 61 7 365 4466. Email: psych.psy.uq.oz.au.

PS 020407

BEST COPY AVAILABLE

Abstract

A theory of working memory in reasoning needs a model of the reasoning processes which utilize that memory, an objective way of assessing the processing loads of reasoning tasks, a model of working memory architecture, and a means of assessing working memory capacity. There is long-standing evidence that much human reasoning is analogical, and recently developed models have shown that analogies can be conceptualized as a mapping from a base structure to a target structure. Previous research has shown that such structure mappings impose a processing load which depends on the amount of information that must be considered jointly in each mapping decision. These processing loads can be assessed objectively by computer simulation of the relevant mapping process. However it is still essential to show empirically that these tasks are capacity-limited, and the easy-to-hard paradigm is probably the best technique for this so far, being free of some of the major limitations of other dual-task paradigms. The literature shows that a unitary model of working memory architecture is untenable, and some tests which are thought to measure general purpose capacity probably measure specialized stores; e.g. memory span probably measures a phonological memory that is specialized for language acquisition. It is difficult to measure capacity of working memory, but there are a number of studies, using widely different methodologies, that suggest it is probably four chunks in adults and three in middle childhood. The lack of adequate measures of working memory capacity has made it impossible to resolve the important question of whether it changes

over age. When the problem is analysed in terms of neural net architectures, the hypothesis that overall capacity remains constant, but that representations become more differentiated with age, appears theoretically and empirically reasonable. This would mean that ability to map structures would increase with age, and would imply that development of reasoning capacity has a maturational component.

Working Memory and Reasoning: The Processing Loads Imposed by Analogies

The capacity and efficienty of working memory are by no means the only factors that are important in cognitive development, but they must be understood before a truly adequate theory can be produced. In this paper I will outline a number of proposals concerning working memory, including its architecture, the reasoning mechanisms that draw on it, ways of assessing task demands and children's capacities, and possible ways working memory might develop with age.

The dominant concept of working memory in cognitive development has been based on short term memory. Following Atkinson and Shiffrin (1968) it has been common to assume that the short term store is the workspace of thinking. However research by Baddeley and Hitch (1974), Halford, Bain and Maybery (1984) and Klapp, Marshburn, and Lester (1983) cast considerable doubt on the view that short term memory is the workspace of thinking. The paradigm used in these studies entails presenting a short term memory preload prior to performing a cognitive task, then requiring recall of the preload after the primary task has been performed. Little interference between the memory preload and the primary cognitive task is found, suggesting that short term memory is not the workspace of thinking. It seems more likely that it is specialised for phonological storage (Baddeley, 1990).

Another possible basis for working memory is primary memory, defined by William James (1890) as information that has not been lost to consciousness. Wickens, Moody and Dow (1981) showed that primary memory

could be distinguished from second memory because the former was not subject to proactive interference whereas the latter was. Halford, Maybery and Bain (1988) used this fact to estimate the capacity of primary memory, and found it to be 4-5 items for adults and 2-3 items for 8-9 year olds.

There is widespread concensus that working memory architecture is more complex than a single memory. Baddeley (1990) has proposed that working memory comprises a central executive and a number of subsidiary systems, including a visuo-spatial scratchpad and an articulatory loop, which is responsible for span tasks. Schneider and Detweiler (1987) postulate a number of regions, each comprising numerous modules. Consistent with contemporary parallel distributed processing (PDP) conceptions of cognition, Schneider and Detweiler postulate no central executive, but do postulate general purpose pathways, of limited capacity, connecting the regions. Their review also leads them to the conclusion that approximately 4 items can be processed in parallel, an estimate that agrees with the independent studies by Halford, Maybery and Bain (1988) mentioned earlier, and by Donald Fisher (1984).

A survey of the working memory literature in the last two decades suggests it is much more complex than a single store. No model of working memory development presented so far is consistent with this literature. This means that the task of modelling working memory development still lies ahead.

One of the most important questions is whether working memory increases or remains constant over age. However much of the evidence that has been claimed to support one side or the other in this debate is inconclusive or even

irrelevant. For example it has been shown that memory span is a function of processing speed (Case, Kurland and Goldberg, 1982; Baddeley, Thomson & Buchanan, 1975) but this does not tell us that processing capacity is constant over age. One reason is that, as we have seen, short term memory span cannot be regarded as a measure of processing capacity. It has also been shown that children who are familiar with a particular content outperform adults who are unfamiliar with it (Chi, Glaser & Rees, 1982). This shows a powerful effect of familiarity on recall, but it does not rule out changes in memory capacity with age. None of these studies provides any measure of capacity, and it is perfectly possible that age differences in capacity were obliterated by large differences in familiarity. Thus much of the claimed evidence that capacity remains constant over age is, at best, inconclusive.

There is a sprinkling of studies in the cognitive development literature suggesting there *are* changes in capacity with age (Halford, Maybery & Bain, 1988; Kail, 1986; Thatcher, Walker & Giudice, 1987; Rudy, in press). See reviews by Fischer (1987) and Halford (1989). However some of these studies have been controversial, and again the evidence is inconclusive.

While acknowledging the prematurity of any firm conclusion about the development of capacity, I would like to suggest that there is not likely to be a single answer to the question. Given the complex nature of working memory architecture as it is now understood, it seems far more likely that capacity for some types of processing remains constant, while others vary to some extent. Incidentally variation with age should not be assumed to mean an increase with

age, even within childhood. For example, there is evidence that recognition of very briefly presented visual stimuli slows down with age (Hoving, Spencer, Robb, & Schulte, 1978). Any attempts to answer questions about the development of capacity will have to take account of the complexity and modularity of working memory architecture. I do not believe we will even begin to solve the capacity problem until we understand better the nature of the capacity we are trying to assess.

Another difficulty has been that the dominant models of cognition had little use for capacity limitations. In models based on the Von Neumann computer metaphor, tasks are segmented into steps that are processed serially, either in a conventional program, or as a production system. Capacity limitations tend not to arise because difficult tasks are broken into more, but smaller steps. Performance depends on availability of the appropriate algorithm, rather than on capacity.

This conception of cognition was interpreted to mean that successful human performance depended on availability of appropriate strategies. Children who failed simply lacked a suitable strategy. A contemporary form of this conception holds that children lack the relevant domain knowledge. Again, capacity limitations are irrelevant in terms of these models. Where capacity has been incorporated into serial processing models it tends to be in the form of a buffer store of limited size. This may have been part of the reason why short term memory, which resembles a buffer in that it holds a small amount of information for later use, was a popular conception of cognitive processing capacity.

By contrast, parallel distributed processing (PDP) models offer a natural interpretation of processing capacity limitations. These models represent information as a set of activation values, normally expressed as a vector, over a set of neural units. Such a vector can represent an arbitrarily large amount of information, but still constitutes a single unit of information. Capacity and resource implications of distributed models have already been discussed in the literature (Anderson, 1973; McClelland, 1986).

New conceptions of capacity are being developed, based on Parallel Distributed Architecture. Halford, Wilson, Guo, Wiles and Stewart (in press) have argued that a vector corresponds to a chunk, in that both comprise a single unit of information of arbitrary size. This offers an explanation for the paradox, which has been with us since Miller's (1956) seminal "magic number" paper, that the limitation is in the number of units, irrespective of the amount of information each contains. If we identify a chunk with a vector, the limitation on the number of chunks implies that we can represent a limited number of independent vectors, irrespective of the amount of information in each.

In order to account for capacity limitations in human inference, a PDP model of analogical reasoning, called the Structural Tensor Analogical Reasoning (STAR) model was developed (Halford et al., in press). This model has been shown by simulation to be capable of handling a variety of analogical reasoning problems. Such a model requires a way to represent relations (more generally, predicates), as well the elements they relate (their arguments). The solution we adopted, shown in Figure 1, is an extension of Smolensky's (1990) approach to the

representation of variables and their values (variable-value bindings).

Insert Figure(s) 1 about here

Figure shows how the fact that an elephant is larger than a dog LARGER_THAN(elephant, dog) is represented in this model. There is one vector representing the relation LARGER_THAN, another representing elephant, and another representing dog. The binding of LARGER_THAN to elephant and dog is represented by the tensor product of the three vectors, shown as the activations of the units within the three-dimensional figure. (For a straightforward account of the mathematics of this representation see Rumelhart & McClelland, 1986, Chapter 9. Details of our formulation are given in Halford et al., in press).

LARGER_THAN is a binary relation, because it has two arguments (a larger and a smaller element). However there are more complex concepts that are based on more than two arguments. Figure 2 shows the representation of a binary operation, such as addition. The figure shows how the fact that 3+2=5 would be represented in this model. One vector represents the ADDITION operation, while the arguments 3,2,5 are each represented by one vector.

Insert Figure(s) 2 about here

Elsewhere (Halford, 1982, 1987, forthcoming) I have argued that the complexity of concepts can be quantified in terms of their dimensionality; that is, the number of independent units of information required to define them. Figure 3 shows how the dimensionality of concepts can be represented in terms of the STAR model.

Insert Figure(s) 3 about here

In general, the number of vectors required to represent a concept is one more than the number of dimensions in the concept, because an additional dimension is required to represent the predicate; that is concepts of dimensionality N are represented by tensor products of rank N+1. Element mappings, which utilise one-dimensional concepts, are represented by the tensor product of 2 vectors. Relational mappings, which utilise two-dimensional concepts, are represented by the tensor product of 3 vectors. System mappings, which utilise three-dimensional concepts, are represented by the tensor product of four vectors, and multiple system mappings, which utilise four-dimensional concepts, are represened by the tensor product of five vectors.

One of the main arguments of my previous research has been that processing loads are greater for concepts of higher dimensionality (Halford, 1987, 1989, forthcoming). The STAR model shows why inferences based on concepts of higher dimensionality become progressively more difficult to compute. The reason is that the size of the tensor product vector increases logarithmically with dimensionality. The essence of the argument is that, if we assume vectors of size M, then concepts of dimensionality N require a tensor product vector of size $M^{(N+1)}$. Furthermore each element in the tensor product vector must be appropriately connected to (at least) one element in each of the component vectors. As the dimensionality of concepts increases, it becomes progressively less likely that the tensor product vectors will have sufficient, appropriate connections to achieve clear representation.

One consequence of the model is that there will be "soft" limits to processing capacity. A requirement for representations of excessively high dimensionality will not lead to catastrophic breakdown in processing, but to unclear representations and ambiguous outputs. The usual result will be a high error rate, rather than a breakdown in processing. Alternately, it will lead to use of representations of lower dimensionality. Some common errors can be predicted from the assumption that representations of inadequate dimensionality are being used.

We will use transitive inference to illustrate the approach. Suppose the premises are: Peter is fairer than Tom, John is fairer than Peter. It is well established that a common way to perform the task is to integrate the premises into

an ordered set, John, Peter, Tom. An analogical reasoning model postulates that this is done by mapping the premises into an appropriate base schema. A suitable schema for this purpose is the "top-bottom" or "left-right" schema that are accustomed to use for arranging things in order. Such a mapping of premises into a schema is shown in Figure 4.

Insert Figure(s) 4 about here

This mapping imposes a processing load because two premises must be processed jointly to determine which element belongs in which position. The premise "Peter is fairer than Tom" only tells us that Peter is in top or middle position. Only by processing the other premise as well, "John is fairer than Peter" can we determine that Peter is in middle position. The processing loads imposed in producing this integrated representation has been confirmed empirically (Maybery, Bain & Halford, 1986; Halford, Maybery & Bain, 1986). Note that the processes used by children here are essentially the same as those used by adults, and the load is high for both.

Insert Figure(s) 5 about here

Figure 5 shows how the transitive inference problem would be represented in the STAR model. There is one vector representing each of John, Peter, and Tom.

There is a vector representing the transitive relation, "fairer than". The four-dimensional structure can be decomposed into three three-dimensional structures, representing "John fairer than Peter", "Peter fairer than Tom" and "John fairer than Tom". The integration of these separate relations into an ordered set, John, Peter, Tom, normally requires the kind of four-dimensional structure shown in Figure 5. The computational difficulty of this representation accounts for the processing load imposed by premise integration.

If the load exceeds capacity, participants default to a simpler representation, using only one relation. For example, if the premises were presented in the order "John is fairer than Peter", and "Peter is fairer than Tom", a child or an adult under high processing load might take account of only the second premise, and answer that Peter is the fairest. There is considerable evidence that children and adults default, on occasions, to processing one premise only, without taking account of earlier premises (Foos, Smith, Sabol & Mynatt, 1976; Maybery, Halford, Bain & Kelly, under review).

Because of the computational demands of multidimensional representations, with consequent high processing loads, it is more efficient to recode concepts into fewer dimensions. In fact, any multidimensional representation can be coded into a single dimension. Thus any of the tensor product representations presented earlier can be recoded as a single vector.

An example would be the transitive inference problem mentioned earlier. The premises "Peter is fairer than Tom", "John is fairer than Peter" can be represented as a rank four tensor product as indicated earlier, or they can be

represented as an ordered triple, "John fairer than Peter fairer than Tom". The latter can be coded as a single vector. I call this process of recoding multidimensional concepts as single dimensional concepts *conceptual chunking*.

The single vector representation obviously makes much less demands on processing capacity than the rank four tensor product. Why then is the more complex tensor product representation ever used?

There are two reasons. First, the structure must be processed before the more efficient representation can be found. In the transitive inference example, the premises must first be recoded into an ordered set before they can be represented as a single vector. To the extent that a structure is new, it cannot be chunked, a fact which is used to test the theory.

Second, the multi-dimensional representation offers greater computational possibilities, and more flexibility. The single vector representation effectively "freezes" the structure, and blocks out other possibilities. Suppose for example, we wanted to consider the alternate premise "Peter is darker than Tom". Computation of the structure entailed by this premise would require the multidimensional representation.

When multidimensional concepts are recoded as single dimensions, they can then be combined with other dimensions without exceeding processing capacity. The concept of velocity is a good example of the way this is done. Velocity is a three dimensional concept, and is defined as distance over time ($V = s/t$). However it can be recoded as a single dimension, which we experience as the position of a pointer on a dial, rush of wind in the face etc.

When velocity is recoded as a single vector in this way, it can be combined with further dimensions to produce higher level concepts. Velocity can be combined with time again to yield acceleration ($a=2st^{-2}$). Acceleration can also be coded as a single dimension, perhaps experienced as the feeling of being thrust back in one' seat. Acceleration can now be combined with mass to yield force, defined as F = ma.

Therefore conceptual chunking, or recoding multidimensional concepts as single dimensions, produces great gains in efficiency, which enables us to bootstrap our way to more complex concepts. The cost is loss of computational possibilities and flexibility. For example, if we express Force as a single vector, the effect of variations in distance and time can no longer be computed.

Another way that efficiency can be improved is by segmentation. That is, complex tasks are broken down into smaller components that are processed serially. Familiarity with a task can lead to strategies that enable concepts to be processed in smaller steps. However the ability to see conceptual systems as a whole may be lost.

We can now ask what develops in this formulation. One major candidate is clearly conceptual chunking. With age and experience the structure of concepts becomes better understood, and they can be recoded into fewer dimensions. The growth of expertise is partly a matter of providing more efficient ways of coding concepts, so they can be treated as conceptual chunks (Chi & Ceci, 1987). Therefore this formulation agrees with Case (1985) to the extent that increased efficiency in the use of available capacity is an important component of

development. However the theories propose diffferent mechanisms. Case's claim (Case 1985, p. 290) that Total Processing Space is a single capacity that can be flexibly allocated to Short Term Storage Space or to Operating Space (TPS = STSS + OS) (Case, 1985, p. 351) is inconsistent with working memory research showing that total processing space is not a unitary store. The alternative that follows from the present theory is that development leads to conceptual chunking, which reduces the number of vectors required to represent a concept. It also leads to strategies that break tasks into smaller steps, thereby reducing the amount of information that has to be processed in parallel at any one time.

Both conceptual chunking and segmentation enable available capacity to be used more efficiently. However the present theory also leads to a reformulation of the question whether capacity changes with age. The way concepts are represented as tensor products of vectors suggests that maturation might lead to differentiation of representations into more vectors. This is consistent with Werner's (1948) theory that differentiation is a major component of development. Differentiation into more vectors would not increase overall capacity, but it would increase the dimensional complexity of concepts that can be represented. This would lead to the ability to process higher level structures.

Increased ability to process structure with maturation would explain why concepts such as transitivity and inclusion, which Piaget insisted were stage-related, are difficult for young children. Empirical work showing that young children have difficulty with these concepts has held up better than is often supposed (Halford, 1989). I have argued elsewhere (Halford, 1987) that they are

both three-dimensional concepts, because both are structures based on three elements (they are ternary relations, or relations with three arguments). Therefore they must be represented as rank four tensor products. My previous research indicates that concepts that require 3-dimensional representations do not develop until approximately 4.5 years. In general, many cases where children have rather intractable difficulties with certain concepts can be attributed to the fact that the concepts entail representations of high dimensionality, which children cannot construct below a certain age.

Conclusions

1. Working memory architecture is complex and modular, with more than one resource. Short term memory span reflects mainly a specialised system, and is not a measure of central processing capacity.

2. Efficiency and capacity of working memory are not the only factors underlying cognitive development, and learning and induction mechanisms are also important.

3. Prior to development of algorithms for a particular domain or situation, human inference is basically analogical, and entails mapping a problem into a base schema.

4. Processing demands of analogical inferences depend on the dimensionality of concepts, defined as the minimum number of elements required to define the concept.

5. The Structural Tensor Analogical Reasoning model, based on PDP (connectionist) architecture, can simulate a range of analogical reasoning processes.

6. In the Structural Tensor Analogical Reasoning model dimensions are represented as vectors, and are identified with chunks.

7. Concepts of dimensionality N are represented as the tensor product of N+1 vectors. Therefore the representation of concepts of high dimensionality entails

computational difficulties, which produce a soft limit on capacity.

8. Transitivity and class inclusion are three-dimensional concepts (ternary relations), which strain the representational capacity of young children.

9. Capacity demands can be reduced by recoding representations into fewer dimensions, known as *conceptual chunking*, or by segmentation, but both lead to lost computational possibilities and less flexibility.

10. Conceptual chunking (more efficient coding), and segmentation (strategy acquisition) underlie cognitive development, but it is possible that with maturation representations become differentiated into more vectors, which does not increase overal capacity, but does increase the dimensionality of structures that can be processed.

REFERENCES

Anderson, J.A. (1973). A theory for the recognition of items from short memorized lists. *Psychological Review, 80*, 417-438.

Atkinson, R.C., & Shiffrin, R.M. (1968). Human memory: A proposed system and its control processes. In K.W. Spence & J.T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory.* (Vol.2, pp. 89-195). New York: Academic Press.

Baddeley, A. (1990). *Human memory: Theory and practice.* Needham Heights, MA: Allyn and Bacon.

Baddeley, A., & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 47-89). New York: Academic Press.

Baddeley, A., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14*, 575-589.

Case, R. (1985). *Intellectual development: Birth to adulthood.* New York: Academic Press.

Case, R., Kurland, M., & Goldberg J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology, 33*, 386-404.

Chi, M.H.T., & Ceci, S.J. (1987). Content knowledge: its role, representation and restructuring in memory development. *Advances in Child Development and Behaviour, 20*, 91-142.

Chi, M.T.H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1, pp. 7-75). Hillsdale, NJ: Erlbaum.

Fischer, K.W. (1987). Relations between brain and cognitive development. *Child Development, 58*, 623-632.

Fisher, D.L. (1984). Central capacity limits in consistent mapping, visual search tasks: four channels or more?. *Cognitive Psychology, 16*(4), 449-484.

Foos, P.W., Smith, K.H., Sabol, M.A., & Mynatt, B.T. (1976). Constructive processes in simple linear order problems. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 759-766.

Halford, G.S. (1982). *The development of thought*. Hillsdale, NJ: Erlbaum.

Halford, G.S. (1987). A structure-mapping approach to cognitive development. *International Journal of Psychology , 22*, 609-642.

Halford, G.S. (1989). Reflections on 25 years of Piagetian cognitive developmental psychology, 1963-1988. *Human Development, 32*, 325-387.

Halford, G.S. (forthcoming). *Children's understanding: the development of mental models*. Hillsdale, N.J.: Erlbaum.

Halford, G.S., Bain, J.D., & Maybery, M.T. (1984). Working memory and representational processes: Implications for cognitive development. In H. Bouma & D.G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 459-470). Hillsdale, NJ: Erlbaum.

Halford, G.S., Maybery, M.T., & Bain, J.D. (1986). Capacity limitations in children's reasoning: A dual task approach. *Child Development, 57*, 616-627.

Halford, G.S., Maybery, M.T., & Bain, J.D. (1988). Set-size effects in primary memory: An age-related capacity limitation?. *Memory and Cognition, 16*(5), 480-487.

Halford, G.S., Wilson, W.H., Guo, J., Wiles, J. and Stewart, J.E.M. (in press). Connectionist implications for processing capacity limitations in analogies. In K. J. Holyoak & J. Barnden (Eds.), *Advances in Connnectionist and Neural Computation Theory, Vol. 2: Analogical Connections* . Norwood, NJ: Ablex.

Hoving, K.L., Spencer, T., Robb, K.Y., & Schulte, D. (1978). Developmental changes in visual information processing. In P.A. Ornstein, *Memory development in children* (pp.21-67). Hillsdale, NJ: Erlbaum.

James, W. (1890). *Principles of psychology*. New York: Holt, Rinehart & Winston.

Kail, R. (1986). Sources of age differences in speed of processing. *Child Development, 57*, 969-987.

Klapp, S.T., Marshburn, E.A., & Lester, P.T. (1983). Short-term memory does not involve the 'working memory' of information processing: The demise of a common assumption. *Journal of Experimental Psychology: General, 112*, 240-264.

Maybery, M.T., Bain, J.D., & Halford, G.S. (1986). Information processing demands of transitive inference. *Journal of Experimental Psychology: Learning, Memory and Cognition, 12*, 600-613.

Maybery, M.T., Halford, G.S., Bain, J.D., & Kelly, M.E. (1988). *Children's construction of series.* Unpublished manuscript, University of Queensland, Brisbane.

McClelland, J.L. (1986). Resource requirements of standard and programmable nets. In Rumelhart, D.E. & McClelland, J.L. (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (pp. 460-487). Cambridge, MA: MIT Press.

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

Rudy, J.W. (in press). Elemental and configural associations, the hippocampus and development. *Developmental Psychobiology*.

Rumelhart, D.E., & McClelland, J.L. (1986). *Parallel Distributed Processing: Explorations in the microstructure of cognition: Vol. 1. Foundations.* Cambridge, MA: MIT Press.

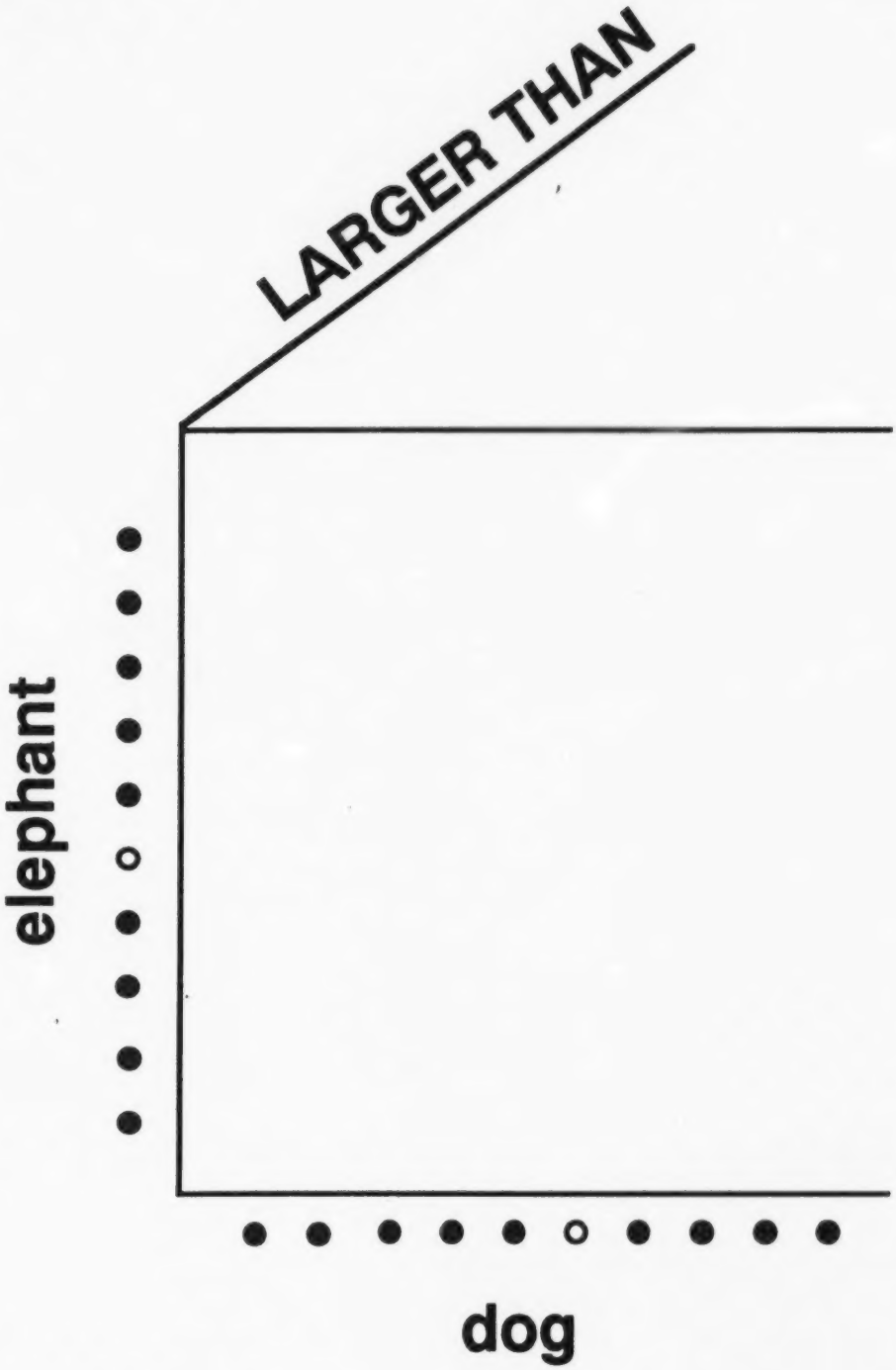
Schneider, W. & Detweiler, M. (1987). A connectionist/control architecture for working memory. *The Psychology of Learning and Motivation, 21*, 53-119.

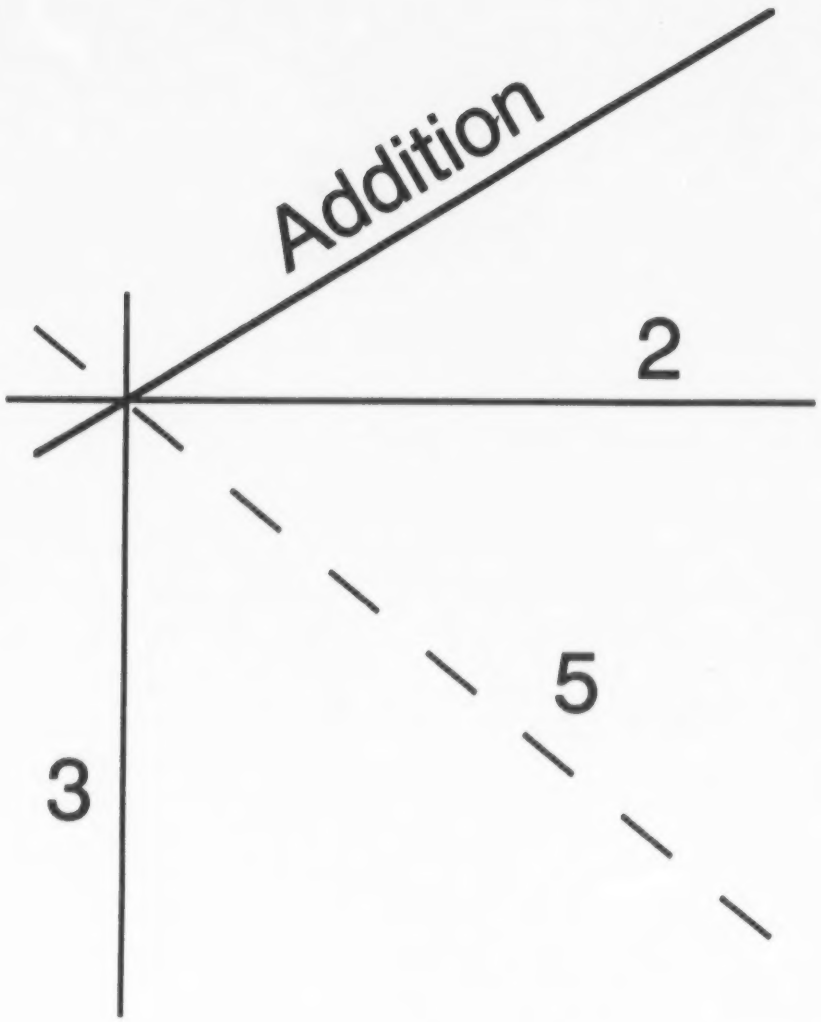
Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence, 46*(1-2), 159-216.

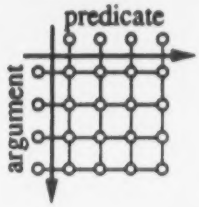
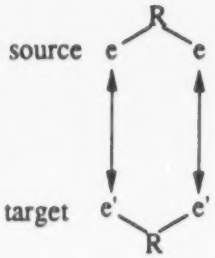
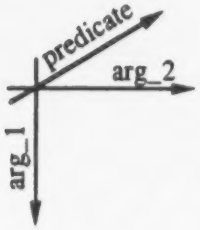
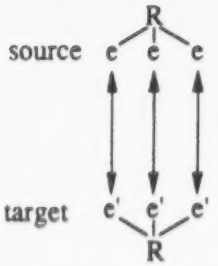
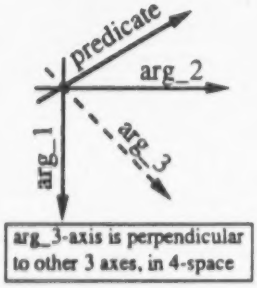
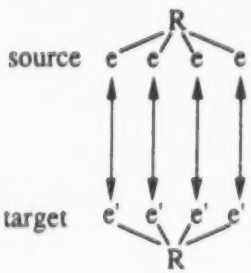
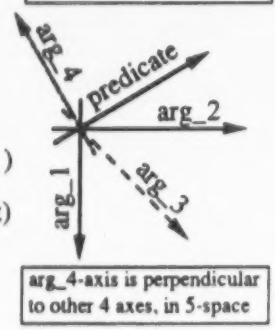
Thatcher, R.W., Walker, R.A., & Giudice, S. (1987). Human cerebral hemispheres develop at different rates and ages. *Science, 236*, 1110-1113.

Werner, H. (1948). *Comparative psychology of mental development*(revised). Chicago: Follet.

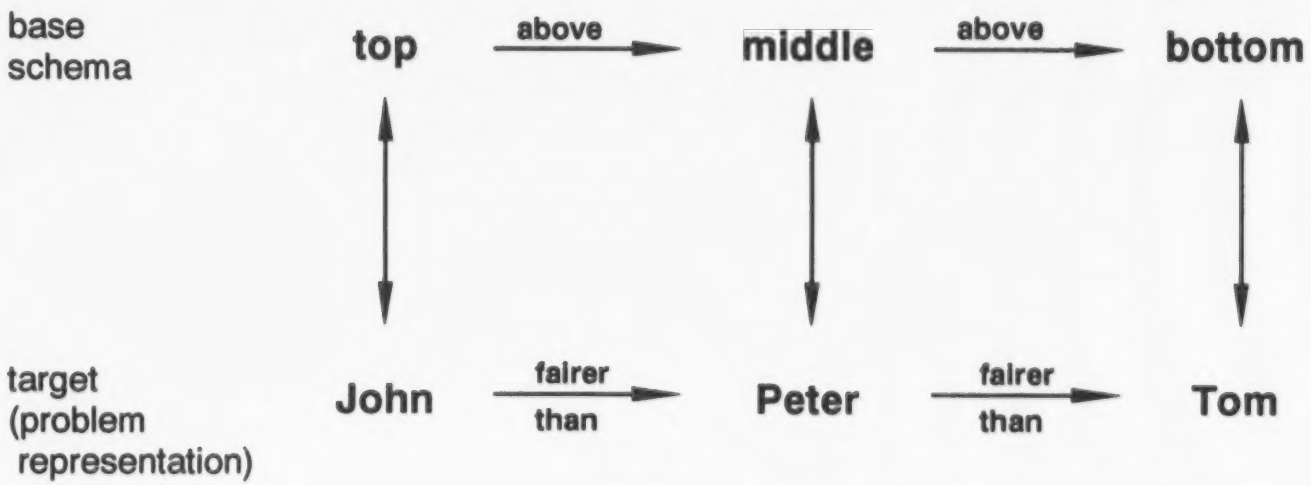
Wickens, D.D., Moody, M.J., & Dow, R. (1981). The nature and timing of the retrieval process and of interference effects. *Journal of Experimental Psychology: General, 110*, 1-20.

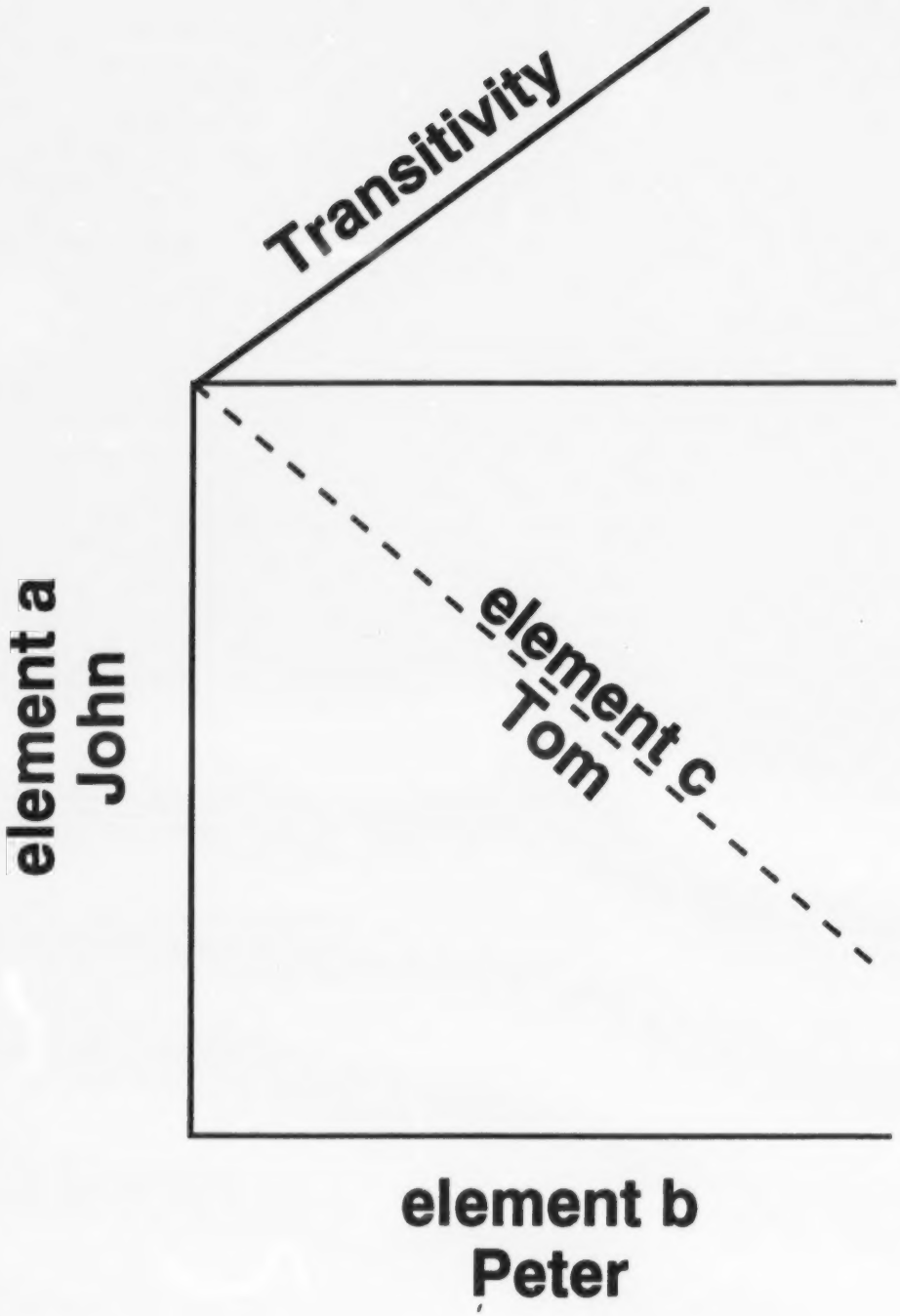

LARGER THAN
elephant
dog


Addition
2
5
3

	Structure Mapping Diagram	Definition	Tensor Product Representation
Element mappings		M: R(e) ⇔ R(e')	
Relational mappings		M: $R(e_i, e_j) \Leftrightarrow R(e'_i, e'_j)$ *e.g.* (a>b) ⇔ (c>d)	
System mappings		M: $R(e_i, e_j, e_k) \Leftrightarrow R(e'_i, e'_j, e'_k)$ *e.g.* (a>b>c) ⇔(r>s>t)	
Multiple system mappings		M: $R(e_i, e_j, e_k, e_l) \Leftrightarrow$ $R(e'_i, e'_j, e'_k, e'_l)$ *e.g.* (a/b=c/d) ⇔ (q/r=s/t)	

Premises: Peter is fairer than Tom
John is fairer than Peter




Transitivity
element a
John
element c
Tom
element b
Peter